ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/51147065

Reservoirs of richness: Least disturbed tropical forests are centres of undescribed species diversity

Article *in* Proceedings of the Royal Society B · May 2011

DOI: 10.1098/rspb.2011.0433 · Source: PubMed

CITATIONS
136

READS
421

6 authors, including:

Xingli Giam
University of Tennessee at Knoxville
123 PUBLICATIONS **4,468** CITATIONS
SEE PROFILE

Brett R Scheffers
University of Florida
110 PUBLICATIONS **11,490** CITATIONS
SEE PROFILE

Gerardo Ceballos
National Autonomous University of Mexico
342 PUBLICATIONS **26,108** CITATIONS
SEE PROFILE

Paul R. Ehrlich
Stanford University
526 PUBLICATIONS **66,426** CITATIONS
SEE PROFILE

All content following this page was uploaded by Gerardo Ceballos on 05 June 2014.

The user has requested enhancement of the downloaded file.

PROCEEDINGS OF THE ROYAL SOCIETY B

Proc. R. Soc. B (2012) **279**, 67–76
doi:10.1098/rspb.2011.0433
Published online 18 May 2011

Reservoirs of richness: least disturbed tropical forests are centres of undescribed species diversity

Xingli Giam[1,*], Brett R. Scheffers[2], Navjot S. Sodhi[2,*], David S. Wilcove[1,3], Gerardo Ceballos[4] and Paul R. Ehrlich[5]

[1]*Department of Ecology and Evolutionary Biology, and* [3]*Woodrow Wilson School of Public and International Affairs, Princeton University, Princeton, NJ 08544, USA*
[2]*Department of Biological Sciences, National University of Singapore, Singapore 117543, Republic of Singapore*
[4]*Instituto de Ecología, Tercer Circuito Exterior S/N, Ciudad Universitaria, Universidad Nacional Autónoma de México, México 04510, Mexico*
[5]*Center for Conservation Biology, Department of Biology, Stanford University, Stanford, CA 94305, USA*

In the last few decades, there has been a remarkable discovery of new species of plants, invertebrates and vertebrates, in what have been called the new age of discovery. However, owing to anthropogenic impacts such as habitat conversion, many of the still unknown species may go extinct before being scientifically documented (i.e. 'crypto-extinctions'). Here, by applying a mathematical model of species descriptions which accounts for taxonomic effort, we show that even after 250 years of taxonomic classification, about 3050 amphibians and at least 160 land mammal species remain to be discovered and described. These values represent, respectively, 33 and 3 per cent of the current species total for amphibians and land mammals. We found that tropical moist forests of the Neotropics, Afrotropics and Indomalaya probably harbour the greatest numbers of undescribed species. Tropical forests with minimal anthropogenic disturbance are predicted to have larger proportions of undescribed species. However, the protected area coverage is low in many of these key biomes. Moreover, undescribed species are likely to be at a greater risk of extinction compared with known species because of small geographical ranges among other factors. By highlighting the key areas of undescribed species diversity, our study provides a starting template to rapidly document these species and protect them through better habitat management.

Keywords: amphibians; conservation; endangerment; generalized linear mixed models; mammals; tropics

1. INTRODUCTION

The start of the new decade brought renewed hope to the Earth's beleaguered biota as we celebrated the United Nations International Year of Biodiversity in 2010. Celebrations and community outreach programmes were held to increase public awareness and the understanding of the importance of biodiversity to humanity and, hopefully, stimulate stronger initiatives among governments, private corporations and individuals to safeguard it. However, efforts to conserve biodiversity are constantly being undermined by anthropogenic activities [1–5]. Biodiversity losses, including the loss of undescribed species [6] and populations [7], are certain to increase as natural habitats are altered for human uses [8,9]. Our inadequate knowledge of biodiversity hampers our understanding of patterns and rates of extinctions, thus undermining our ability to develop effective strategies to conserve it [10]. Moreover, undescribed species could be important for ecosystem processes and the delivery of ecosystem services [11]. These reasons alone underscore the importance of discovering today's unknown species and understanding their ecological significance.

Previous research has evaluated the completeness of species inventories in particular regions [12–14] or globally [9,15–17], and there have been systematic efforts to guide species discovery (e.g. The Census of Marine Life [18]). However, the distribution of undescribed species across regions and taxa, the parameters that may influence the likelihood of finding new species and the degree to which undescribed species may be threatened by extinction are all poorly understood. This is despite the fact that knowledge of these factors could inform policy decisions such as the allocation of resources to species discovery and the expansion of protected area networks to guard ecosystems estimated to harbour rich undescribed biotas.

Using two taxonomic groups, amphibians and land mammals, we attempt to predict the distribution of undescribed species across biomes nested in biogeographic realms (hereafter, realm-biome units, e.g. Neotropical tropical moist forests). Mammals and amphibians are relatively well studied, as exemplified by the recent completion of species conservation assessments in both groups [19,20]. The assessments reveal that roughly 30 and 21 per cent of all described amphibian and mammal species, respectively, are endangered. Mapping

* Authors for correspondence (xgiam@princeton.edu; dbsns@nus.edu.sg).

Electronic supplementary material is available at http://dx.doi.org/10.1098/rspb.2011.0433 or via http://rspb.royalsocietypublishing.org.

Received 25 February 2011
Accepted 27 April 2011

This journal is © 2011 The Royal Society

distributions across realm-biomes—previously used in identifying key areas of future environmental change [21]—is appropriate for our analyses because they are distinct biogeographic regions comprising of unique ecosystems and species assemblages [21] and harbour a sufficiently large number of species to allow the detection of broad and reliable patterns. We recorded all amphibian and land mammal species, their year of description and Geographic Information Systems (GIS) range maps from the International Union for Conservation of Nature and Natural Resources (IUCN) Global Amphibian Assessment [18] and the IUCN Global Mammal Assessment [19] (see §2). We address the following questions: (i) Are there still relatively large numbers of undescribed amphibian and land mammal species? (ii) Which regions probably contain the largest numbers of undescribed species, and are such regions congruent between land mammals and amphibians? (iii) Are recently discovered species more likely to be endangered compared with those long known to science? (iv) Can biome type and the magnitude of human impact predict the proportion of undescribed species within each region? and (v) Does the current extent of the protected area network correlate with the projected number of undescribed species across regions?

2. METHODS

(a) *Datasets*

We assembled a list of all extant and extinct amphibian ($n = 6296$) and land mammal species ($n = 5398$), their year of description and their 2010 IUCN threat status from the IUCN Red List website (www.iucnredlist.org). The IUCN Red List dataset does not include species which went extinct before 1500. For both amphibians and mammals, the first species was described in 1758 and the most recent species in 2010. We downloaded range maps of amphibian and land mammal species from the Global Amphibian Assessment and Global Mammal Assessment on the IUCN Red List website. Range maps for 6188 of 6296 (98.2%) amphibian species and 5264 of 5398 (97.5%) land mammal species are available. Species that were not mapped are presented in the electronic supplementary material, table S1 (amphibians) and table S2 (mammals).

We delimited terrestrial biomes nested in biogeographic realms (realm-biomes) following the biogeography of the World Wildlife Fund (WWF) Ecoregions of the World dataset [22]. After excluding the Antarctic biogeographic realm as there was no native amphibian or land mammal species, 63 realm-biome units across seven biogeographic realms remained for further analyses (electronic supplementary material, table S3). We overlaid range maps of native amphibian and land mammal species onto the WWF Ecoregions of the World dataset to generate a species list for each realm-biome unit using ArcGIS v. 9.3 software [23]. We defined a species to be native to a particular realm-biome only if that realm-biome overlaps with more than 5 per cent of the species' total range size. This is necessary to minimize false-positives—a species is mistakenly attributed to a particular realm-biome owing to a small, but erroneous overlap between its range and the realm-biome area—as the IUCN range maps have a relatively coarse resolution. Native species lists in mangrove realm-biomes were generated by extracting mangrove species from the IUCN Red List website and overlaying them with maps of each biogeographic realm. This step is important for obtaining accurate mangrove species lists as mangrove areas are small and we are likely to obtain an erroneous species list if we simply overlay the range maps of all species with the mangrove area [24,25].

We calculated the mean human footprint [26] for every realm-biome by overlaying the Human Footprint Index v. 2 (available at http://sedac.ciesin.columbia.edu/wildareas) dataset onto the WWF Ecoregions of the World dataset. This human footprint index, expressed as a percentage of the relative local human impact in each terrestrial biome, is an aggregate index of anthropogenic activity synthesized from human population density, land transformation, access and electric power infrastructure datasets.

(b) *Number and proportion of undescribed species within each realm-biome unit and across the globe*

After generating native species lists for each of the 63 realm-biomes, we estimate the total number of undescribed amphibian and land mammal species in each realm-biome using a mathematical model which incorporates the change in taxonomic effort over time. This model was proposed by Joppa *et al.* [9] and Pimm *et al.* [14], and it was previously applied to estimate the number of undescribed species in global angiosperms, as well as angiosperms and major vertebrate groups in Brazil.

First, the expected number of species described in a 5 year period t, $\hat{S}_t$ is given by some fraction, k_t, of the total number of undescribed species up to the same 5 year period t, U_t,

$$\hat{S}_t = k_t(U_t). \tag{2.1}$$

The number of undescribed species in a 5 year period t, U_t, can be calculated by subtracting the cumulative number of species described up to the previous 5 year period $t-1$, N_{t-1}, from total number of species, N_{tot},

$$U_t = N_{\text{tot}} - N_{t-1}. \tag{2.2}$$

The fraction k_t can be thought of as the probability of describing one species from the remaining pool of undescribed species. Unlike previous studies [13,16] that did not take into account the taxonomic effort, the fraction, k_t, in equation (2.1) is not a constant because an increase in taxonomic effort is likely to increase the probability of finding and describing a species in any particular 5 year time period t. We express k_t as a product of taxonomic effort Tax_t, and an efficiency variable E_t,

$$k_t = Tax_t E_t. \tag{2.3}$$

The taxonomic effort, Tax_t, is given by the number of first-author and second-author taxonomists actively describing species in the time period t in each biogeographic realm. We define the active working period of each taxonomist as the years spanning the author's first and last species descriptions as the first- or second-author. We chose to consider only up to two taxonomists per species because species described by two authors were not uncommon in early taxonomy (e.g. S. Müller & Schlegel, Peters & Doria in the 1800s, Robinson & Kloss, Tate & Archibold in early 1900s), while the recent increase in the number of authors in the molecular taxonomy era is likely to reflect greater collaboration rather than increased per-author taxonomic effort. In a few cases, there were no available working taxonomists in a few given pre-1800 time periods. This is not surprising given that

early taxonomists tended to publish multiple species descriptions in a single monograph in a single given year. To avoid underestimating the taxonomic effort in these early time periods, they were assigned with the number of taxonomists working in the preceding period.

Following Joppa *et al.* [9] and Pimm *et al.* [14], we assume that the efficiency in finding species, E_t, increases linearly over time, owing to technological advancements such as the introduction of air travel over the last century,

$$E_t = a + bt. \tag{2.4}$$

The 5 year time period, t, is scaled such that the first time period is 0, second one is 1, third one is 2, etc. As E_t represents the probability of finding species per taxonomist, it has to be a positive value. We therefore set a lower-bound for parameters a and b at 0 to ensure that the optimization routine (discussed later) always returns a positive value for E_t.

By substituting equations (2.2), (2.3) and (2.4) into (2.1), we obtain a mathematical equation that allows us to estimate the number of species described in any given 5 year time period, S_t,

$$\hat{S}_t = Tax_t(a + bt)(N_{\text{tot}} - N_{t-1}). \tag{2.5}$$

The number of species described in each time period tends to be spiky in the earlier years of species discovery (*ca* before 1900) owing to the publication of species monographs which report an accumulation of species discovered in years prior to their publication, resulting in residuals that are non-normal as well as a non-homogeneous variance. This pattern was also reported for global angiosperms [9], as well as angiosperms and vertebrates in Brazil [14]. To obtain unbiased estimates of the unknown parameters a, b, and N_{tot}, we first normalized the residuals and stabilized the variance by square-root transforming both the observed number of species, S_t, as well as the predicted number of species, $\hat{S}_t$ which is equal to $Tax_t(a + bt)(N_{\text{tot}} - N_{t-1})$ (equation (2.5)). Following that, we minimized the sum of squares of their differences (residuals) by performing constrained optimization using PORT routines [27] implemented by the nlminb function in R v. 2.12 [28]. The variable N_{tot} was constrained to be no less than the total number of species described to date. Realm-biomes with less than 20 described species were excluded from the analysis because of low sample size. We recognize that the number of species in each time period may still be low in some realm-biomes but this is unlikely to bias our results as model plots were checked for non-normal residuals and heteroscedasticity. Subsequent analysis of the model results (via logit GLMMs) gives larger weights to realm-biomes with large n, hence, thus allowing our conclusions to be relatively robust to small sample sizes in some realm-biomes.

Unlike previous studies [9,14], we started the first 5 year period from 1758 (i.e. first time period 1758–1762, second time period 1763–1767) because species described by Linneaus in the monograph *Systema Naturae* 10th edition [29] did not appear to make the time period 1758–1762 an outlier for most realm-biomes. We excluded the most recent time period 2008–2012 from the analysis because not all of the most recently described species have been included in the IUCN database. Therefore, the most recent time period considered in this study is 2003–2007. We recognized that the large number of species described by Linneaus in *Systema Naturae* as well as by other taxonomists in monographs published in later time periods may be outliers and hence bias our estimates in a few realm-biomes. We therefore adapted a statistical approach based on false discovery rates [30] to identify outlying time periods for each realm-biome (refer to electronic supplementary material, appendix 1 for the detailed procedure). After removing outliers we re-ran the optimization routine as presented above and obtained an estimate for the total number of species N_{tot} for each realm-biome. For each realm-biome, we visually inspected the square-root transformed plot of fitted versus observed values of the number of species ($\hat{S}_t$ and S_t, respectively) to assess model fit. We also checked the assumption of a linear relationship between efficiency and year (equation (2.4)) by comparing the square-root transformed fitted $(a + bt)$ and observed values $(=S_t/[(N_{\text{tot}} - N_{t-1})Tax_t])$ Only a single data point (amphibians in Neotropical temperate broadleaf and mixed forests: code NT4) was excluded as efficiency appeared to decrease with year. These plots are presented in the electronic supplementary material (appendices 2 and 3).

The number of undescribed species in each realm-biome N_{undesc} is calculated by subtracting the total number of described species N_{desc} from the total number of species estimated from the model in equation (2.5),

$$N_{\text{undesc}} = N_{\text{tot}} - N_{\text{desc}}. \tag{2.6}$$

The number of undescribed species expressed as a proportion of the estimated total number of species (hereafter, proportion of undescribed species P_{undesc}) is therefore,

$$P_{\text{undesc}} = \frac{N_{\text{undesc}}}{N_{\text{tot}}}. \tag{2.7}$$

To evaluate the certainty of our estimates of undescribed species, we used a delete-one observation jackknife resampling procedure where we generated all possible resamples of the data after removing a single 5 year time period [9,14]. We then re-ran the model using these resamples and reported the lower and upper quartile of the estimates of the number and proportion of undescribed species.

Lastly, we re-ran the model for the global dataset of amphibians and land mammals to estimate the total number and proportion of undescribed species in these two taxonomic groups.

(c) *Correlation between the year of description and extinction proneness*

We fitted generalized linear mixed-effect models (GLMMs) using the lmer function implemented in R v. 2.12 to investigate whether the year of description of a species was correlated with the probability of a species being threatened. We did so to assess whether species described more recently are likely to stand a greater chance of being imperilled. If this is true, and by extrapolating the trend, undescribed species are more likely to be imperilled compared with species already known to science.

We followed IUCN's convention whereby species classified as critically endangered (CR), endangered (EN) and vulnerable (VU) are considered threatened species. We excluded species classified by IUCN as data-deficient (DD) because their conservation status could not be determined owing to a lack of data. As species are phylogenetic units with shared evolutionary histories, they are not statistically independent [31,32]. To account for this phylogenetic effect, we coded the random-effects structure of the

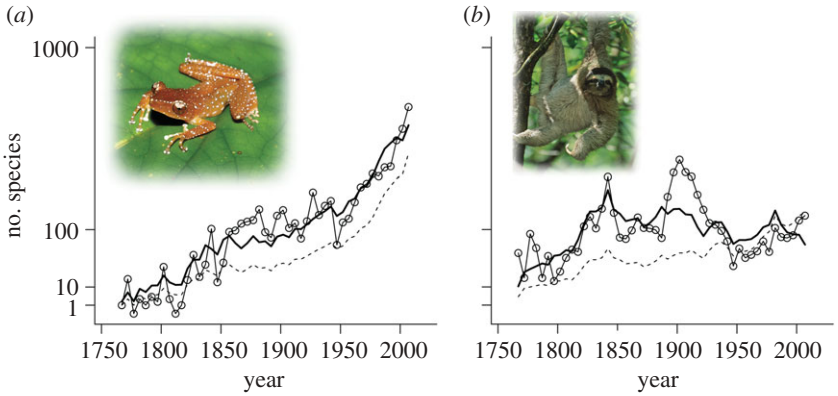

Figure 1. Species description curves of (*a*) amphibians and (*b*) land mammals showing the actual number of described species (solid line with hollow circles), modelled number of described species (bold line), and number of taxonomists (dashed line) in each 5 year period up to 2007. The amphibian shown is the spotted tree frog (*Nyctixalus pictus*) endemic to the tropical moist forests of Southeast Asia. The mammal is the critically endangered pygmy three-toed sloth (*Bradypus pygmaeus*), restricted to Isla Escudo de Veraguas in the islands of Bocas del Tora, off the Caribbean coast of Panama.

GLMM as a taxonomic effect [31]. We followed the protocol suggested by Zuur *et al.* [33] to determine the random-effects structure for each taxonomic group. For the amphibians, the intercept coefficient was allowed to vary across Orders, while for land mammals, both the slope and intercept coefficients were allowed to vary across Orders. All GLMMs had a binomial response variable (threatened or not) and a logit link. We used an information-theoretic measure–Akaike's Information Criterion corrected for small sample sizes (AIC_c) to assess the relative strengths of evidence of the competing candidate models [34]. The relative support of each model being the best of the candidate model set was calculated using AIC_c weights ($wAIC_c$) with the weights varying from 0 (no support) to 1 (complete support). To evaluate the amount of variance in the response variable that is explained by each model, we calculated the deviance explained by each model (%DE) relative to the null model with the random intercept but no fixed effects [31].

(d) *Correlation between the relative proportion of undescribed species with anthropogenic disturbance and biome type*

To investigate whether tropical forest biomes and human footprint predict the proportion of undescribed species across realm-biomes, we fitted five candidate binomial GLMMs with logit link to the data using the lmer function implemented in R v. 2.12 [28]. Each GLMM is a hypothesized relationship between one or both of these predictor variables and the estimated proportion of undescribed species within each realm-biome (electronic supplementary material, table S4). In each GLMM, biogeographic realm was entered as a random intercept to account for the shared geographical history and taxonomic sampling patterns between realm-biomes within each continent. Similar to the previous analysis, we used $wAIC_c$ to evaluate support and %DE to calculate variance explained for each model.

(e) *Concordance between the proportion of undescribed species and protected area network*

We collated data on the current extent of protected areas in each realm-biome from Jenkins & Joppa [35]. We tested the concordance between the proportion of undescribed amphibian and land mammal species and the proportion of land area gazetted as protected areas across realm-biomes using Kendall rank correlation implemented in R v. 2.12.

3. RESULTS

(a) *Species discovery trends*

There has been a continuous increase in the number of amphibian species described in each successive 5 year period since the description of the first species in 1758 (figure 1*a*). In contrast, in mammals there was an increase in the number of species described in each 5 year period until about 1850, after which the number of new species has remained relatively constant (figure 1*b*). For both taxonomic groups there was an increase in the number of first-author taxonomists actively working in each 5 year period. Overall, our model predicted a global amphibian species richness of 9347 [interquartile range (IQR): 9276–9511] and global land mammal species richness of 5561 (IQR: 5557–5571). The number of undescribed amphibian species is predicted to be roughly 3051 (32.6% of total predicted species richness), while that of undescribed land mammal species is predicted to be 163 (2.9% of total predicted species richness).

(b) *Number of undescribed biodiversity across realm-biomes*

We applied the species description model to obtain the estimated number of undescribed species in each individual realm-biome containing 20 or more described species. As an example, we demonstrate the results of our model for land mammals across tropical moist broadleaf forests of Australasia, the Afrotropics, Indomalaya, the Neotropics and the Palearctic (figure 2). After an initial increase in species descriptions until *ca* 1850, the number of new species described per 5 year interval remained relatively constant for moist forests in Australasia, the Afrotropics, Indomalaya and the Neotropics. A small upswing in the number of species can be observed in the Afrotropical, Indomalayan, and Neotropical moist forests after 1950, probably due to the increase in taxonomic effort (dashed lines in figure 2*b*,*c*,*d*). In the Palearctic moist forests, species descriptions have declined since the initial increase and no new species were described since 1950, suggesting that few, if any, undescribed species exists in the Paleartic moist forests. As seen from figure 2, our model provided reasonably good fits to the general trends in species descriptions.

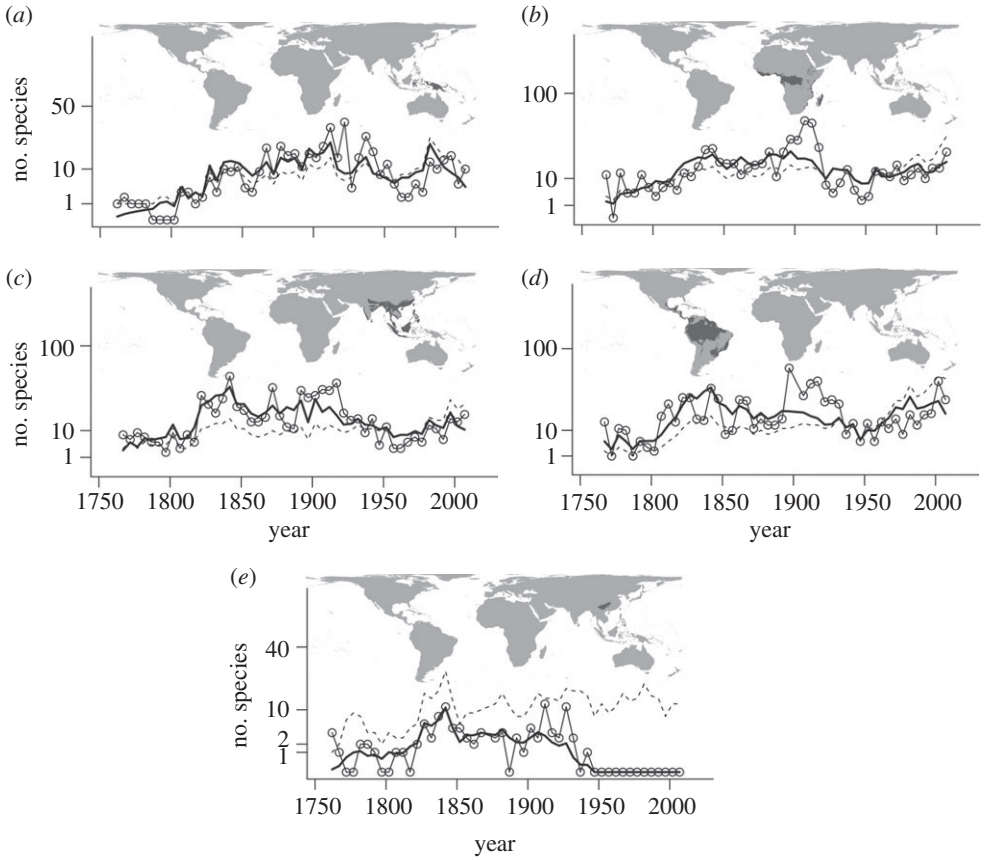

Figure 2. Species description curves of land mammals in tropical moist forest realm-biome units. (*a*) Australasia; (*b*) Afrotropics; (*c*) Indomalaya; (*d*) Neotropics; (*e*) Palearctic. In each panel, the number of described species (solid line with hollow circles), modelled number of described species (bold line), and number of taxonomists (dashed line) in each 5 year period up to 2007 are presented.

The estimated number and proportion of undescribed species are not homogeneous across realm-biome units (figure 3). The Neotropical moist broadleaf forests were predicted to harbour the largest number of undescribed amphibian species (1781; IQR: 1770–1973) and land mammal species (80; IQR: 76–85). The proportion of undescribed amphibian species was the highest in the Australian moist broadleaf forests (56.3%; IQR: 55.3–61.1%) while that of land mammals was highest in the Palearctic Mediterranean forests and woodlands (12.0%; IQR: 11.0–12.0%). Summation of the predicted number of undescribed species across realm-biome yielded 3506 undescribed amphibian species, and 410 undescribed mammal species.

The estimated number of undescribed amphibian species is positively correlated with that of undescribed land mammal species across realm-biomes (Kendall rank correlation coefficient $\tau = 0.27$, $p = 0.02$). However, the proportion of undescribed amphibians was not correlated with the proportion of undescribed land mammals across realm-biomes (Kendall $\tau = 0.13$, $p = 0.28$). For each taxonomic group, we present the five realm-biomes with the highest predicted undescribed species richness (table 1). Results for all realm-biomes are available in the electronic supplementary material (table S5).

(c) *Extinction risk in recently described species*

The likelihood of a species being listed as threatened on the 2010 IUCN Red List increases with the year of its scientific description (amphibians, $n = 4661$, figure 4*a*; land mammals, $n = 4530$, figure 4*b*). This relationship is consistent across all three taxonomic orders within the amphibians and all 25 taxonomic orders within the land mammals with the exception of Perissodactyla (odd-toed ungulates), where probability of endangerment is relatively constant across the year of description. The year of description of a species explains 9.8 and 8.4 per cent of the deviance in the binary response of whether an amphibian or land mammal species is threatened or not, respectively. For both taxa, this GLMM provides an approximately 100 per cent better fit to the data compared with the stochastic null model as determined by sample size corrected Aikaike's Information Criterion weights—$w\text{AIC}_c$ (table 2). It is plausible that the more recently described species are more likely to be listed as threatened because they are not well studied and hence their population or range size is underestimated. New species discovered owing to taxonomic splitting are likely to have a smaller range size. To investigate whether our results are an artefact of the autocorrelation between date of discovery and range size, we reanalysed the data after removing species that are listed as threatened owing to a limited range and/or population size (IUCN threat criteria B and D), leaving only species threatened owing to an observed or projected decrease in population size (amphibians, $n = 3075$; land mammals, $n = 3905$). The relationship between extinction proneness and year of description persists—species described more recently are more likely to be threatened (electronic

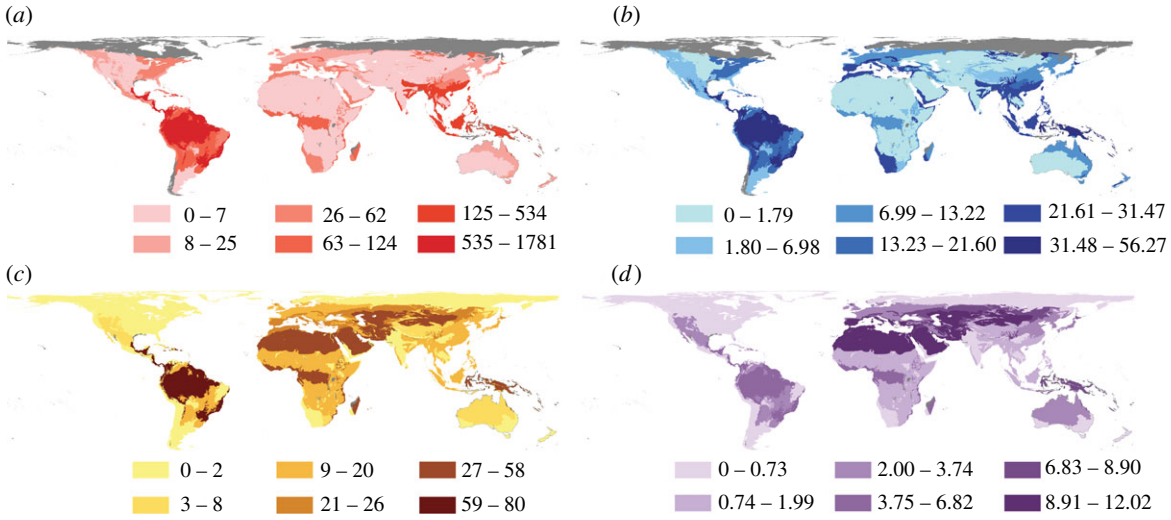

Figure 3. The distribution of undescribed amphibian and land mammal species across global biome-realms. Shown are the number of undescribed (*a*) amphibian and (*c*) mammal species, and the percentage of undescribed (*b*) amphibian and (*d*) mammal species, as predicted from our model. Estimates are not available for the biome-realms coloured grey owing to a low sample size ($n < 20$) or the lack of model convergence. The maps have a cylindrical equal-area projection.

Table 1. The top five realm-biomes ranked in order of estimated undescribed species richness. N_{desc}, number of described species currently known to science; N_{tot}, estimated total number of species; N_{undesc}, estimated number of undescribed species; %, proportion of undescribed species in each realm-biome, expressed as a percentage. IQR, the interquartile range for the estimated parameters. Full results are presented in the electronic supplementary material, table S5.

realm-biomes	N_{desc}	N_{tot} (IQR)	N_{undesc} (IQR)	%
all amphibians	9347	9347 (9276–9511)	3051 (2980–3215)	32.64
Neotropics—tropical and subtropical moist broadleaf forests	4454	4454 (4443–4646)	1781 (1770–1973)	39.99
Australasia—tropical and subtropical moist broadleaf forests	949	949 (929–1068)	534 (514–653)	56.27
Indomalaya—tropical and subtropical moist broadleaf forests	1340	1340 (1327–1353)	404 (391–417)	30.15
Neotropics—Montane grasslands and shrublands	394	394 (379–397)	124 (109–127)	31.47
Afrotropics—tropical and subtropical moist broadleaf forests	846	846 (844–853)	95 (93–102)	11.23
all land mammals	5398	5561 (5557–5571)	163 (159–173)	2.93
Neotropics—tropical and subtropical moist broadleaf forests	1173	1253 (1249–1258)	80 (76–85)	6.38
Afrotropics—tropical and subtropical moist broadleaf forests	821	879 (879–882)	58 (58–61)	6.60
Australasia—tropical and subtropical moist broadleaf forests	471	517 (516–518)	46 (45–47)	8.90
Indomalaya—tropical and subtropical moist broadleaf forests	888	906 (905–907)	18 (17–19)	1.99
Neotropics—tropical and subtropical grasslands, savannas and shrublands	560	580 (580–581)	20 (20–21)	3.45

supplementary material, figure S1, table S6). While the amount of deviance explained by the model is reduced to 3.1 and 2.1 per cent for amphibians and land mammals, respectively, the GLMM that predicts extinction proneness from year of description is still 100% better than the null model.

(d) *Tropical forests, human footprint and the distribution of undescribed biodiversity*

The two GLMMs containing human footprint index and biome type were consistently the two most parsimonious models as ranked by AIC_c (table 3). These two models explain up to 55 and 44 per cent of the deviance in amphibians and land mammals, respectively. These most parsimonious models demonstrate that the proportion of undescribed species is higher in tropical forests compared with other biome types (figure 5). The relationship between human footprint index and the proportion of undescribed species is dependent on the biome as well as the taxa. In tropical forest biomes, the proportion of both undescribed amphibians and land mammals increases strongly with decreasing human footprint (figure 5). In non-tropical forest biomes, the proportion of undescribed amphibians increases weakly with decreasing human footprint while that of mammals remain relatively constant.

(e) *Concordance between the proportion of undescribed species and protected area network*

Existing protected area coverage [35] is correlated with neither the proportion of undescribed land mammal species (Kendall $\tau = -0.02$, $p = 0.88$) nor that of amphibians across realm-biomes (Kendall $\tau = 0.15$, $p = 0.17$). Protected area coverage is generally low in many habitats where the proportion of undescribed species is likely to be large (e.g. Afrotropical dry broadleaf forests where only 6% of its area is protected [35]).

4. DISCUSSION

Our model, when applied to the global dataset, indicates that a large number of species (approx. 3050) remain

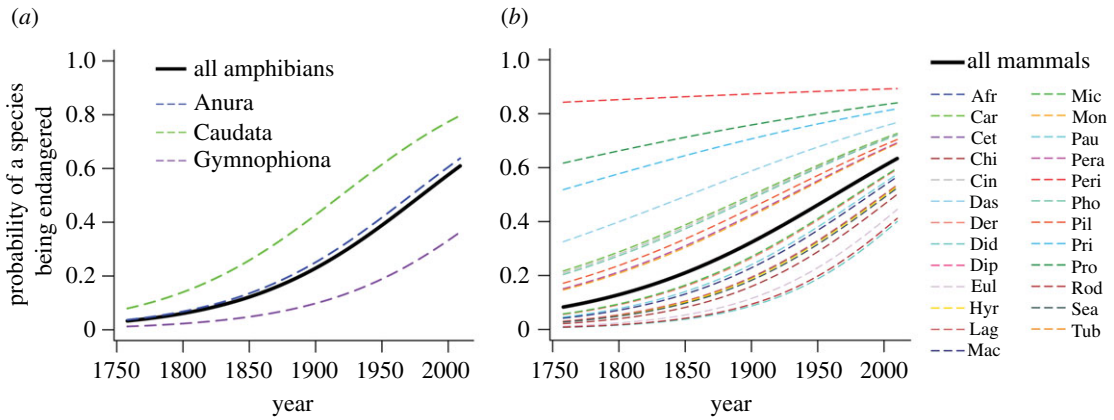

Figure 4. Best-fit GLMMs demonstrating that extinction proneness increases with the year of description for both amphibians (*a*) and land mammals (*b*) (see table 2 for multimodel inference statistics). Black lines represent population level fitted values. The fitted relationship between year of description and proportion of species threatened within each phylogenetic (order) group are represented by colored, dotted lines. Mammal order abbreviations—Afr: Afroscoricida, Car: Carnivora, Cet: Cetartiodactyla, Chr: Chiroptera, Cin: Cingulata, Das: Dasyuromorphia, Der: Dermoptera, Did: Didelphimorphia, Dip: Diprotodontia, Eul: Eulipotyphla, Hyr: Hyracoidea, Lag: Lagomorpha, Mac: Macroscelidea, Mic: Microbiotheria, Mon: Monotremata, Pau: Paucituberculata, Pera: Peramelemorphia, Per: Perissodactyla, Pho: Pholidota, Pil: Pilosa, Pri: Primates, Pro: Proboscidea, Rod: Rodentia, Sca: Scandentia, Tub: Tubulidentata.

Table 2. Generalized linear mixed models (GLMMs) investigating the relationship between year of description and probability of endangerment for (*a*) amphibians and (*b*) mammals. The models are ranked by Akaike's Information Criterion corrected for small sample size (AIC_c). Predictor terms shown are year = year of description. Also shown are the number of parameters (*k*), log likelihood (*LL*), the difference in AIC_c of each model from the highest ranked model (ΔAIC_c), AIC_c weights representing the probability of each model being the best ($wAIC_c$), and the per cent deviance explained by each model (%DE).

rank	model	*k*	*LL*	AIC_c	ΔAIC_c	$wAIC_c$	%DE
(*a*) amphibians ($n = 4661$, 1898 threatened)							
1	~year	3	−2816.71	5639.43	0	≈1	9.82
2	null	2	−3123.56	6251.13	611.70	≈0	
(*b*) land mammals ($n = 4530$, 1142 threatened)							
1	~year	4	−2198.07	4404.16	0	≈1	8.43
2	null	2	−2400.42	4804.85	400.69	≈0	

undescribed. For land mammals, the number is smaller but still sizeable (approx. 160). For amphibians, the summation of regional estimates (approx. 3500 species) supported the estimate derived from the global dataset while for land mammals, the sum of the number of undescribed species across regions (approx. 400 species) was more than two times higher than that of the global dataset. This suggests that there is greater uncertainty in the expected number of undescribed mammal species and that it may be slightly underestimated in the global dataset. The distribution of undescribed species is likely to be heterogenous—some biomes have more undescribed species than others. The proportion of undescribed species within each realm-biome also varies across the globe. The weak correlation between the estimated number of undescribed amphibian and land mammal species suggests that the spatial trends in undescribed species richness may differ among taxonomic groups. However, this hypothesis can only be tested when global range maps of other taxa become available.

Tropical forests are likely to have the largest proportion of their species undescribed as they are structurally complex [36] and contain more species compared with any other biome [37]. Moreover, tropical forests are less well-studied and well-explored compared with other biomes. The proportion of undescribed species increases with decreasing human footprint in tropical forest biomes. This could be owing to the fact that tropical forests with low human footprint are less well-explored compared with areas with a higher level of human impact. It is also possible that human disturbance increases the rate of crypto-extinctions of undescribed species, therefore the pool of extant undescribed species, as predicted by species discovery trends, is larger in tropical forests with low human footprint. New species described from splitting cryptic allopatric species due to recent surveys in areas that were previously free from human influence, and therefore underexplored, may partially contribute to the high undescribed species richness in realm-biomes of low human footprint. On the other hand, the relationship between the proportion of undescribed species and human footprint is weak in non-tropical forest biomes. A possible reason for this is that non-tropical forest biomes are less dense structurally compared with tropical forest biomes [36] and therefore likely to have similar accessibility across the human footprint gradient. The heterogeneous distribution of undescribed biodiversity suggests that conservation

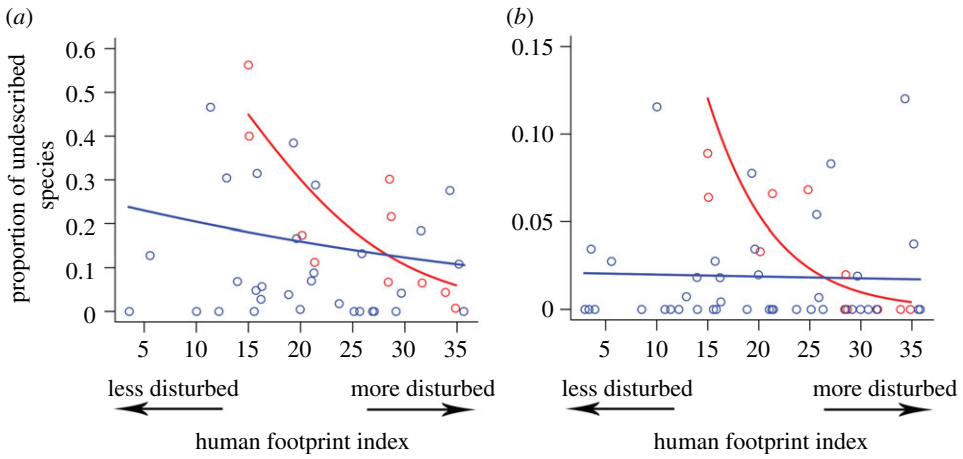

Figure 5. Best-fit GLMMs demonstrating the correlation between proportion of undescribed species and human footprint and biome type (see table 3 for multimodel inference statistics). Tropical forests with minimal anthropogenic disturbance are likely to contain the highest proportion of undescribed species. Red dots represent tropical forest realm-biomes while blue dots represent other realm-biomes. Red lines illustrate the relationship between human footprint index and proportion of newly described species in tropical forests while blue lines illustrate the same relationship in other biomes.

Table 3. Generalized linear mixed models (GLMMs) investigating the predictors of the proportion of undescribed (*a*) amphibian species and (*b*) mammals species ranked by AIC_c. Predictor terms shown are HF = human footprint index and TropFor = tropical forest biome or not. Also shown are the number of parameters (*k*), log likelihood (*LL*), the difference in AIC_c of each model from the highest ranked model (ΔAIC_c), AIC_c weights representing the probability of each model being the best ($wAIC_c$), and the per cent deviance explained by each model (%DE).

rank	model	*k*	*LL*	AIC_c	ΔAIC_c	$wAIC_c$	%DE
(*a*) amphibians ($n = 42$ realm-biomes)							
1	~HF + TropFor + HF*TropFor	5	−332.39	676.49	0.00	≈1	54.62
2	~HF + TropFor	4	−402.76	814.63	138.14	≈0	45.01
3	~HF	3	−557.36	1121.37	444.88	≈0	23.90
4	~TropFor	3	−582.66	1171.98	495.49	≈0	20.44
5	null (intercept only)	2	−732.38	1469.08	792.59	≈0	
(*b*) land mammals ($n = 52$ realm-biomes)							
1	~HF + TropFor + HF*TropFor	5	−103.48	218.26	0.00	≈1	43.50
2	~HF + TropFor	4	−143.45	295.74	77.49	≈0	21.68
3	~TropFor	3	−153.80	314.11	95.85	≈0	16.02
4	~HF	3	−177.49	361.48	143.23	≈0	3.09
5	null (intercept only)	2	−103.48	218.26	0.00	≈0	

policy can be designed to target biomes likely to contain a large proportion of undescribed species, specifically, least-disturbed tropical forests.

We showed that extinction-proneness of amphibian and land mammal species increases with year of description, i.e. recently described species are more likely to be endangered compared with species known to science for a long time. By extrapolating this trend, many of the undescribed species in these two groups are probably in danger of extinction and could well disappear before they are discovered, especially given the high rates of habitat conversion in tropical forests. This is consistent with predictions that a large proportion of today's undescribed species may go extinct without ever being recorded [10,38]. Some of these cryptic biodiversity losses could have impacts on ecosystem functioning and ecosystem services [10,11]. The relationship between extinction proneness and year of description persists even when range-restricted species are removed; suggesting that extinction proneness is mediated by genuine population declines of recently described species in addition to small range/population sizes and is not solely an artefact of the latter. However, owing to the decrease in explained deviance in the model after removing range-restricted species, we can infer that a small range size also drives endangerment in recently described species. It is possible that the population/range size of recently described species is underestimated owing to incomplete knowledge on their distribution. If the population/range size estimates are accurate, future searches for new species should operate at small spatial scales [39]. For example, it may be necessary to set up multiple closely spaced sampling sites to avoid missing undescribed species with small ranges.

The estimated number and proportion of undescribed species are not correlated with the existing protected area coverage. An expansion in the protected area network is most urgently required in biomes where the proportion of undescribed species is high but the current extent of protected area is low. There is great potential for expanding protected areas (but doing so in a way that respects

the interests and values of indigenous peoples) through innovative international cooperation. In May 2010, for example, Indonesia placed a two-year moratorium on the establishment of new forest concessions for logging and agricultural conversion under a deal in which Norway will invest $1 billion in forest conservation and capacity building for future Reduced Emissions from Deforestation and Degradation (REDD) projects [40]. For long-term biotic conservation, donor countries should negotiate for the establishment of protected areas in recipient countries. Complementary to protected areas are other approaches such as conservation easement programmes, which may help to protect key habitats from development while offsetting opportunity costs incurred by private landowners [41], and ecosystem services programmes that add economic and human welfare arguments [42] in support of conserving key habitats for undescribed biodiversity.

The 'perfect storm' for biodiversity loss is upon us. Universities and funding agencies have weakened their support for taxonomic exploration and research [43], with little prospect of reversing this trend. Meanwhile, the very biomes that probably harbour the greatest number of undescribed species are being altered in ways that will drive many of these species to extinction. Existing templates such as Biodiversity Hotspots [44] and Last of the Wild [26] may be used to prioritize habitat conservation, but they do not explicitly account for the probable distribution of undescribed species. We therefore provide a starting template to efficiently guide species discovery efforts and to increase conservation efforts in those areas likely to harbour large numbers of undescribed yet imperilled species. Novel initiatives such as engaging parataxonomists—local resident biodiversity specialists who have received on-the-job training in classification [45]—or the use of DNA barcoding, combined with further advances in molecular taxonomy [46], may expedite the discovery of unknown species while reducing the costs of biotic exploration. However, in the short term at least, it is imperative to preserve the key habitats of undescribed biodiversity, i.e. relatively undisturbed tropical forests, through better habitat use and management, while fulfilling human development goals. Today's 'hidden' biodiversity need not vanish without a trace. It is up to us to try to prevent such a tragedy.

We thank Trevor Price, Stuart Pimm and an anonymous referee for their insightful comments which greatly improved the paper. We thank Norman Lim for contributing the photo of the spotted tree frog and Bill Hatcher for the pygmy sloth photo. X.G. was supported by the Princeton University Graduate Fellowship. N.S.S. was supported by the National University of Singapore (Grant no. R-154-000-331-112).

REFERENCES

1 Hansen, M. C., Stehman, S. V. & Potapov, P. V. 2010 Quantification of global gross forest cover loss. *Proc. Natl Acad. Sci. USA* **107**, 8650–8655. (doi:10.1073/pnas.0912668107)
2 Thomas, C. D. *et al.* 2004 Extinction risk from climate change. *Nature* **427**, 145–148. (doi:10.1038/nature02121)
3 Clavero, M. & Garcia-Berthou, E. 2005 Invasive species are a leading cause of animal extinctions. *Trends Ecol. Evol.* **20**, 110. (doi:10.1016/j.tree.2005.01.003)
4 Rosser, A. M. & Mainka, S. A. 2002 Overexploitation and species extinctions. *Conserv. Biol.* **16**, 584–586. (doi:10.1046/j.1523-1739.2002.01635.x)
5 Dudgeon, D. *et al.* 2006 Freshwater biodiversity: importance, threats, status and conservation challenges. *Biol. Rev.* **81**, 163–182. (doi:10.1017/S1464793105006950)
6 Stork, N. E. 1993 How many species are there? *Biodivers. Conserv.* **2**, 215–232. (doi:10.1007/BF00056669)
7 Hughes, J. B., Daily, G. C. & Ehrlich, P. R. 1997 Population diversity: its extent and extinction. *Science* **278**, 689–692. (doi:10.1126/science.278.5338.689)
8 Ehrlich, P. R. & Pringle, R. M. 2008 Where does biodiversity go from here? A grim business-as-usual forecast and a hopeful portfolio of partial solutions. *Proc. Natl Acad. Sci. USA* **105**, 11 579–11 586. (doi:10.1073/pnas.0801911105)
9 Joppa, L. N., Roberts, D. L. & Pimm, S. L. 2010 How many species of flowering plants are there? *Proc. R. Soc. B* **278**, 554–559. (doi:10.1098/rspb.2010.1004)
10 Ceballos, G. & Ehrlich, P. R. 2009 Discoveries of new mammal species and their implications for conservation and ecosystem services. *Proc. Natl Acad. Sci. USA* **106**, 3841–3846. (doi:10.1073/pnas.0812419106)
11 Suzán, G., Marcé, E., Giermakowski, J. T., Mills, J. N., Ceballos, G., Ostfeld, R. S., Armién, B., Pascale, J. M. & Yates, T. L. 2009 Experimental evidence for reduced rodent diversity causing increased hantavirus prevalence. *PLoS ONE* **4**, e5461. (doi:10.1371/journal.pone.0005461)
12 Baselga, A. & Novoa, F. 2006 Diversity of Chrysomelidae (Coleoptera) in Galicia, Northwest Spain: estimating the completeness of the region inventory. *Biodivers. Conserv.* **15**, 205–230.
13 Giam, X., Ng, T. H., Yap, V. B. & Tan, H. T. W. 2010 The extent of undiscovered species in Southeast Asia. *Biodivers. Conserv.* **19**, 943–954. (doi:10.1007/s10531-010-9792-2)
14 Pimm, S. L., Jenkins, C. N., Joppa, L., Roberts, D. L. & Russell, G. J. 2010 How many endangered species remain to be discovered in Brazil. *Natureza & Conservação* **8**, 71–77. (doi:10.4322/natcon.00801011)
15 Medellin, R. A. & Soberon, J. 1999 Prediction of mammal diversity on four land masses. *Conserv. Biol.* **13**, 143–149. (doi:10.1046/j.1523-1739.1999.97315.x)
16 Bebber, D. P., Marriott, F. H. C., Gaston, K. J., Harris, S. A. & Scotland, R. W. 2007 Predicting unknown species numbers using discovery curves. *Proc. R. Soc. B* **274**, 1651–1658. (doi:10.1098/rspb.2007.0464)
17 Hamilton, A. J. *et al.* 2010 Quantifying uncertainty in estimation of tropical arthropod species richness. *Am. Nat.* **176**, 90–95. (doi:10.1086/652998)
18 Yarincik, K. & O'Dor, R. 2005 The census of marine life: goals, scope and strategy. *Sci. Mar.* **69**(Suppl. 1), 201–208.
19 IUCN 2010 IUCN Global Amphibian Assessment. See www.iucnredlist.org. Accessed Feb 2010.
20 Schipper, J. *et al.* 2008 The status of the world's land and marine mammals: diversity, threat, and knowledge. *Science* **322**, 225–230. (doi:10.1126/science.1165115)
21 Lee, T. M. & Jetz, W. 2008 Future battlegrounds for conservation under global change. *Proc. R. Soc. B* **275**, 1261–1270. (doi:10.1098/rspb.2007.1732)
22 Olson, D. M. *et al.* 2001 Terrestrial ecoregions of the world: A new map of life on Earth. *BioScience* **51**, 933–938. (doi:10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2)
23 ESRI 1999–2008 *ArcGIS 9.3*.
24 Lamoreux, J. F., Morrison, J. C., Ricketts, T. H., Olson, D. M., Dinerstein, E. ric, McKnight, M. W. & Shugart, H. H. 2006 Global tests of biodiversity concordance

and the importance of endemism. *Nature* **440**, 212–214. (doi:10.1038/nature04291)
25 Guilhaumon, F., Gimenez, O., Gaston, K. J. & Mouillot, D. 2008 Taxonomic and regional uncertainty in relationships and the identification of richness hotspots. *Proc. Natl Acad. Sci. USA* **105**, 15 458–15 463. (doi:10.1073/pnas.0803610105)
26 Sanderson, E. W., Jaiteh, M., Levy, M. A. & Redford, K. H. 2002 The human footprint and the last of the wild. *Bioscience* **52**, 891–904. (doi:10.1641/0006-3568(2002)052[0891:THFATL]2.0.CO;2)
27 Gay, D. M. 1990 Usage summary for selected optimization routines. Computing Science Technical Report No. 153. Murray Hill, NJ, AT & T Bell Laboratories..
28 R Development Core Team 2010 *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
29 Linneaus, C. 1758 *Systema Naturae*, 10th edn. Stockholm, Sweden: Laurentius Salvius.
30 Motulsky, H. J. & Brown, R. E. 2006 Detecting outliers when fitting data with nonlinear regression—a new method based on robust nonlinear regression and the false discovery rate. *BMC Bioinformatics* **7**, 1–20. (doi:10.1186/1471-2105-7-123)
31 Bradshaw, C. J. A., Giam, X., Tan, H. T. W., Brook, B. W. & Sodhi, N. S. 2008 Threat and invasive status in legumes is related to opposite extremes of the same ecological and life-history attributes. *J. Ecol.* **96**, 869–883. (doi:10.1111/j.1365-2745.2008.01408.x)
32 Felsenstein, J. 1985 Phylogenies and the comparative method. *Am. Nat.* **125**, 1–15. (doi:10.1086/284325)
33 Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A. & Smith, G. M. 2009 *Mixed effect models and extensions in ecology with R*. New York, NJ: Springer.
34 Burnham, K. P. & Anderson, D. R. 2002 *Model selection and multimodel inference: a practical information-theoretic approach*, 2nd edn. New York, NY: Springer.
35 Jenkins, C. & Joppa, L. N. 2009 Expansion of the global protected area system. *Biol. Conserv.* **142**, 2166–2174. (doi:10.1016/j.biocon.2009.04.016)
36 Whitmore, T. C. 1984 *Tropical rain forests for the Far East*, 2nd edn. Oxford, UK: Oxford University Press.
37 Bradshaw, C. J. A., Sodhi, N. S. S. & Brook, B. W. 2009 Tropical turmoil: a biodiversity crisis in progress. *Front. Ecol. Environ.* **7**, 79–87. (doi:10.1890/070193)
38 Gaston, K. J. & Spicer, J. I. 2004 *Biodiversity: an introduction* 2nd edn. Oxford: Wiley-Blackwell.
39 Levin, S. A. 1992 The problem of pattern and scale in ecology. *Ecology* **73**, 1943–1967. (doi:10.2307/1941447)
40 Clements, G. R., Sayer, K., Boedhihartono, A. K., Venter, O., Lovejoy, T., Koh, L. P. & Laurance, W. F. 2010 Cautious optimism over Norway–Indonesia REDD pact. *Conserv. Biol.* **24**, 1437–1438. (doi:10.1111/j.1523-1739.2010.01584.x)
41 Gustanski, J. A. & Squires, R. H. 2000 *Protecting the land: conservation easements past, present, and future*. Washington, DC: Island Press.
42 Daily, G. C. *et al.* The value of nature and the nature of value. *Science* **289**, 395–396. (doi:10.1126/science.289.5478.395)
43 Kozlowski, G. 2008 Is the global conservation status assessment of a threatened taxon a utopia? *Biodivers. Conserv.* **17**, 445–448. (doi:10.1007/s10531-007-9278-z)
44 Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B. & Kent, J. 2000 Biodiversity hotspots for conservation priorities. *Nature* **403**, 853–858. (doi:10.1038/35002501)
45 Janzen, D. H. 2004 Setting up tropical biodiversity for conservation through non-damaging use: participation by parataxonomists. *J. Appl. Ecol.* **41**, 181–187. (doi:10.1111/j.1365-2664.2004.00879.x)
46 Smith, M. A., Rodriguez, J. J., Whitfield, J. B., Deans, A. R., Janzen, D. H., Hallwachs, W. & Hebert, P. D. N. 2008 Extreme diversity of tropical parasitoid wasps exposed by iterative integration of natural history, DNA barcoding, morphology, and collections. *Proc. Natl Acad. Sci. USA* **105**, 12 359–12 364. (doi:10.1073/pnas.0805319105)

View publication stats